DOCUMENT RESUME

ED 231 848 TM 830 364

AUTHOR Nagel, Stuart S.
TITLE Ethical Dilemmas in Policy/Program Evaluation.
PUB DATE Oct 82
NOTE 31p.; Paper presented at the Annual Meeting of the Evaluation Research Society (Baltimore, MD, October 1982).
PUB TYPE Speeches/Conference Papers (150) -- Reports - Research/Technical (143)

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS *Ethics; Evaluation Methods; *Evaluation Needs; Evaluation Utilization; Evaluators; *Policy Formation; *Program Evaluation; *Research Problems; Standards; *User Satisfaction (Information)

ABSTRACT

In this context, ethics refers to a set of normative standards for resolving dilemmas that policy analysts or evaluators often face with regard to what they should do in performing their roles in the interests of societal desirability. The dilemmas relate to: (1) prediction versus prediction plus prescription; (2) evaluation versus evaluation plus diverse replication; (3) whether or not to work to maximize the interests of a political party/special interest group; (4) focusing on intended consequences versus all consequences; (5) efficiency versus equity; (6) cost incurring versus cost saving in research; (7) whether or not to share one's raw data-sets; (8) research validity versus questionable findings that are not sufficiently questioned; and (9) whether or not to put people at risk in policy evaluation. On the basis of this analysis one can possibly arrive at an overall conclusion that policy evaluators operating on a high level of professional responsibility should seek to develop optimum policies and to do optimum research. Optimum policies refer to policies that maximize societal benefits minus societal costs. Optimum research refers to research that maximizes beneficial new knowledge minus the cost of obtaining it. (PN)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

✓ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Stuart Nagel

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

ED231848

ETHICAL DILEMMAS IN POLICY/PROGRAM EVALUATION

Stuart S. Nagel, University of Illinois

TM 830 369

For Presentation on the 1982 ERS-ENet Panel on
"Ethical Dilemmas in Evaluation Research and Practice"
October, 1982, Draft

Abstract of

ETHICAL DILEMMAS IN POLICY EVALUATION

Stuart S. Nagel, University of Illinois

In this context, ethics refers to a set of normative standards for resolving dilemmas that policy analysts or evaluators often face with regard to what they should do in performing their roles in the interests of societal desirability. The dilemmas relate to (1) prediction versus prediction plus prescription, (2) evaluation versus evaluation plus diverse replication, (3) whether or not to work to maximize the interests of a political party/special interest group, (4) focusing on intended consequences versus all consequences, (5) efficiency versus equity, (6) cost incurring versus cost saving in research, (7) whether or not to share one's raw data-sets, (8) research validity versus questionable findings that are not sufficiently questioned, and (9) whether or not to put people at risk in policy evaluation.

Openness may be the key factor underlying the resolving of ethical dilemmas in policy evaluation. The situation is like blue sky laws in regulating the securities markets. A stockbroker can legitimately sell the blue sky so long as he informs potential customers and others that is what they are getting. Likewise, policy evaluators may be complying with basic ethical obligations when they make clear exactly what they are doing on matters like the above nine dilemmas. Perhaps, however, more affirmative action may be needed for resolving these ethical dilemmas than just openness in reporting one's research. In the analogy to the blue sky laws, one might argue that a stockbroker should have an ethical obligation not to sell the blue sky, rather than merely an obligation to report that she or he is selling the blue sky. On the above nine dilemmas in policy analysis research, the more affirmative position might involve the following:

1. Policy analysts should seek to find the policy or combinations of policies that will maximize a given goal or goals.
2. They should always do a sensitivity analysis to determine how their prescriptions might differ as a result of changing the measurement, entities, assumptions, or other inputs of their basic analysis.
3. They should not aid certain groups if the goals of the groups are illegal, such as racist redistricting designed to decrease the influence of minority groups below a reasonably proportionate level.
4. They should check for side effects and unforeseen consequences in doing policy analysis.
5. They should be concerned with equity as a minimum constraint, and also as a constraint that is subject to a sensitivity analysis for determining how the prescriptions would differ if the equity constraints were changed.
6. They should try to keep research costs down.
7. They should seek to facilitate the secondary re-use of data which they have gathered.
8. They should be careful to avoid findings of questionable validity, including failure to control for alternative explanations or failure to use proper measurements, samples, and assumptions.
9. Policy analysts should be careful not to put people at risk, or else be sure informed consent is involved if people are put at risk, although policymakers can and should be subject to being exposed as wrongdoers or as having undesirable biases in their decision-making.

On the basis of this analysis one can possibly arrive at an overall conclusion that policy evaluators operating on a high level of professional responsibility should seek to develop optimum policies and to do optimum research. Optimum policies refer to policies that maximize societal benefits minus societal costs. Optimum research refers to research that maximizes beneficial new knowledge minus the costs of obtaining it. Those are high and often impossible goals to achieve. By seeking them, however, policy evaluation is likely to achieve more than it would with any lesser set of goals.

ETHICAL DILEMMAS IN POLICY EVALUATION

The purpose of this article is to suggest and discuss some principles of ethics in policy evaluation. In this context, ethics refers to a set of normative standards for resolving dilemmas that policy analysts or evaluators often face with regard to what they should do in performing their roles in the interests of societal desirability. Those standards, as discussed in this article, include situations that involve policy optimizing, sensitivity analysis, partisanship, unforeseen consequences, equity, efficient research, research sharing, research validity, and handling official wrongdoing. Each dilemma situation will be discussed in terms of some general ideas and a concrete illustrative example. The examples will often draw on the field of criminal justice research, but one can easily reason by analogy to other subject matters.

The dilemmas discussed are nine in number. They are grouped in terms of ethical dilemmas that relate to appropriate purposes/goals/effects and those that relate to appropriate methods. Within the first group are dilemmas that relate to (1) prediction versus prediction plus prescription, (2) whether or not to work to maximize the interests of a political party/special interest group, (3) focusing on intended consequences versus all consequences, (4) efficiency versus equity as policy goals. Within the methods group are dilemmas that relate to (5) evaluation versus evaluation plus diverse replication, (6) cost incurring versus cost saving in research, (7) whether or not to share one's raw data-sets, (8) research validity versus questionable findings that are not sufficiently questioned, and (9) whether or not to put people at risk in policy evaluation.[1]

I. ETHICAL DILEMMAS THAT RELATE TO APPROPRIATE PURPOSES, GOALS, AND EFFECTS

A. THE OPTIMIZING DILEMMA

Policy evaluators, analysts, and researchers often present articles, reports, and findings that deal with the relations between policies and goals. Merely presenting those relations can often cause policy makers, the general public, or other researchers to jump to false conclusions as to what policies should be adopted. To what extent does a policy evaluator have an obligation to clarify the relations between (1) predictive or causal findings and (2) policy decisions as to what ought to be done about a given problem? In

other words, have policy evaluators performed their roles in a socially desirable way by abstaining from discussing the policy implications of their findings? This is the dilemma of deciding whether one should just be concerned with prediction and causation, or should one also be concerned with prescription and possibly policy optimization.

To be more specific, policy researchers in the criminal justice field are often interested in what variables correlate with (1) having less crime per capita across places, or (2) having reduced crime per capita over time. James Q. Wilson, for example, in the widely read Family Weekly of November 15, 1981, says "Evidence on the effect of punishment suggests that certainty is more important than severity in deterring criminals." That kind of statement has been made by many crime-policy researchers. It is a statement that implies society should be devoting more of its resources to catching and convicting criminals than to imprisoning them for lengthy periods.

Suppose it is true that in a statistical analysis, one can predict crime rates better from knowing the percentage of crimes that result in arrests or convictions than one can from knowing the sentence lengths of those who are convicted. It does not, therefore, follow that society should do more or spend more trying to catch and convict criminals and do less with regard to length of imprisonment. Implicit or explicit statements about the relative value of increasing certainty versus increasing severity are meaningless unless one especially considers (1) how high the probability of getting caught already is, (2) how high the severity of the penalty already is, (3) how high the benefits from crime-committing are, (4) how steep the relations are among crime occurrence, certainty, severity, and crime benefits, and (5) how much the monetary and other costs are for incremental increases in certainty and severity.

On the fifth matter alone, a little thought can indicate how meaningless an impact or effects study can be for directly drawing policy conclusions without considering the benefits and costs that are associated with those impacts. For example, suppose a 1-unit increase in certainty does produce a 10-unit decrease in crime, whereas a 1-unit increase in severity may only produce a 5-unit decrease in crime. It does not, therefore, follow that we should concentrate our marginal or other resources on increasing certainty because a 1-unit increase in certainty may cost more than twice as much as a 1-unit increase

in severity, and increasing certainty may produce undesirable side effects. That, however, depends on how the increase in certainty is obtained.

The important point is that policy evaluators may be undesirably misleading people by not discussing the policy implications of their findings. That does not mean the researcher has to do an elaborate analysis of what it would cost to increase certainty and severity. It may, however, mean that the researcher should at least raise the issue that concentrating resources on relatively strong policies may be unwise from a benefit-cost perspective. In this particular context, the researcher might also have an obligation to note (1) the diminishing returns of concentrating on a given policy, (2) the possibility of developing an alternative policy that was not part of the original analysis, such as reducing crime benefits , and (3) the interactive relations among policies such that certainty of being convicted means nothing if there are no substantial penalties associated with being convicted as a result of a lack of prison resources.[2]

B. THE PARTISAN DILEMMA

In doing policy evaluation, the evaluator may sometimes be called upon to determine the effects of alternative policies on highly partisan goals such as maximizing the number of legislative districts dominated by Democrats or dominated by Republicans. Likewise, the evaluator may be called upon to aid in prescribing a policy that will maximize the interests of some special interest group such as Blacks, union members, or automobile manufacturers. Under those circumstances what should the ethical policy evaluator do, particularly in light of what is good for society?

One can argue that a policy evaluator ought to be able to give advice to any legal group for achieving any legal purpose through legal means. Lawyers can give advice, even to illegal groups, although they cannot ethically advise clients to do illegal things. A justification for encouraging or allowing policy evaluators to give partisan advice is that if each side has good advisors, then the truth and best policies should be more capable of winning out in the interplay of both sides seeking to present their cases as well as they can be presented. That is a key argument

in favor of the adversary legal system whereby lawyers go all out for each side and a better judicial decision is thereby facilitated. Evaluators, though, should not be considered to be too much like lawyers. There are at least two differences on an ethical dimension. One is that lawyers have no obligation to qualify their statements, but ethical evaluators (even when working for partisan causes) should indicate how their conclusions are dependent on their inputs, as previously mentioned. A second difference is that lawyers not only advise but also advocate. Policy analysts should probably mainly advise and generally leave the advocacy to others. That is especially true of argumentative public speaking, as contrasted to analysts advocating a policy in a prescriptive or optimizing mode to their clients or in the policy literature.

The redistricting field is one that is particularly good for illustrating partisan dilemmas. In the 1960s when computerized redistricting was first being developed, there was substantial conflict among the original researchers concerning the propriety of including political party information as input into the redistricting models. The purist position was to leave it out entirely to avoid corruption of the nonpartisan evaluations. A compromise position was to input the information so the districting patterns which were outputted could show how many Democrats and how many Republicans there would be in each district since that information is useful to policy makers and other interested persons. A position that went further involved programming the computer to draw district lines that would provide equal population, and also provide a percentage of districts dominated by the Democrats equal to the percentage of Democratic voters in the state. That type of evaluation could be justified as providing a form of bipartisan proportional representation. A position that went even further involved programming the computer to draw district lines to maximize the number of Democratic-dominated districts or Republican-dominated districts. That capability was justified on the grounds that by showing the Democrats and the Republicans how far each could go (given the constraints of nearly equal population per district and party registration figures), the redistricting evaluation system thereby facilitated moving toward a compromise between those two extreme positions.

The same capability that could maximize Democratic-dominated districts could also be used to maximize black-dominated or white dominated districts, which raises related

but more complicated problems. Trying to maximize white-dominated districts may be clearly illegal, especially if it means no black representation at all as it could. Trying to maximize black-dominated districts is not illegal since Blacks are still likely to be a minority in any state legislature. Maximizing black-dominated districts may, however, be contrary to the interests of Blacks in seeking favorable state or city legislation. For example, in the state of Mississippi and the city of Chicago, Blacks constitute about 40% of the population. They would perhaps be better off if the district lines were drawn in such a way that they would dominate zero districts, but constitute 40% of every district. That could be done in Chicago by making each district look like a pie slice with the center of the pie in the heart of the black population. That kind of line drawing would make Blacks the swing vote between the white candidates in every district and thereby guarantee that close to 100% of the elected candidates will be highly favorable to black interests. That kind of line drawing, however, is likely to be opposed by Blacks who are incumbent or would-be legislators since they tend to want safe black-dominated districts. The policy evaluator in that context would possibly have a dilemma between the interests of the black population and the black leaders. Evaluators who work for political organizations are likely to wind up drawing maps which favor the political leaders, whereas more academic evaluators might draw politically unfeasible maps which favor general population categories. Both kinds of evaluators, though, are performing their roles in an acceptable manner by at least pointing out those possibilities.[3]

C. THE UNFORESEEN CONSEQUENCES DILEMMA

Policy evaluators should be concerned with trying to foresee what might otherwise be the important unforeseen consequences of the policies with which they are working. The failure to foresee important consequences may sometimes constitute a form of negligence that would amount to evaluation malpractice. Someday evaluators may regularly take out malpractice insurance to protect themselves from paying damages for the harm their lack of foresight may cause. It is better, though, to be foresightful than to be insured. Just because a consequence is unforeseen does not mean that it is unforeseeable.

The ability to foresee the effects of alternative public policies can be substantially improved if one has good predictive models of the behavior of the people one is

interested in. For example, if one is trying to foresee the effects of alternative criminal justice policies, one had better have a model of the criminal justice system that includes plea bargaining since it (rather than trials) is the essence of the criminal justice system. The relevance of that kind of model to foreseeing unforeseen consequences can be illustrated by the attempt of the Vera Foundation of New York City to reduce the pretrial jail population (-Y) by increasing pretrial release (X_1). At first glance, one might think such a goal would obviously be improved by such a policy. What seems obvious, however, may not be so. By increasing pretrial release, one thereby decreases the number of defendants who are sitting in jail awaiting trial. Doing so makes them substantially less vulnerable to pleading guilty in return for the prosecutor's offer to recommend probation or a sentence equal to the time they have already served in jail awaiting trial. The detained defendant might argue that he is innocent and that he wants a trial to prove his innocence. The prosecutor can inform him that he is indeed entitled to a trial, but a trial is not likely to be held for a few months, and he can go home now if he will plead guilty. That is an offer which is difficult to refuse, especially if the defendant's witnesses may not be so reliable, if he has been previously convicted and thus will not greatly further damage his resume, and if he knows that upon conviction he may serve a substantial prison sentence.

If fewer defendants are now in jail (X_1), that means fewer guilty pleas (X_2) and probably more trials. If there are more trials (X_3), that means more delay in the system to process cases. If there is more delay in the system (X_4), then that means defendants sitting in jail awaiting trial will wait even longer than they were waiting before the increase in pretrial release. If the waiting time in jail goes up, then the jail population may also go up (+Y) since the jail population, like the world population, is dependent on how many births or inputs there are and how long people live or stay in the system. If longevity goes up, it may more than offset a decrease in births or inputs. Then the total population in the world or in the jails increases rather than decreases, even though there has been a decrease in inputs into the jail as a result of the increased pretrial release. This could be an especially undesirable unforeseen consequence because the policy designed to reduce the jail population actually increases it. Most unforeseen consequences

involve adverse side effects, rather than adverse effects on the key goal being sought.

A dilemma occurs in the sense that the evaluator, on the one hand, cannot be paralyzed out of fear of unforeseen consequences. We do want evaluators to be willing to go forward and evaluate alternative public policies so we have more rational policy-adoption than we would have without policy evaluation. On the other hand, evaluators should make a strong effort to try to develop contingency models that show what is likely to happen if certain contingent events occur. Such models can be useful in planning what to do to take into consideration a variety of consequences. In the above example, a contingency plan could have been developed whereby prosecutors sweeten their offers of shorter sentences to those who are released so as to hold constant the number of defendants pleading guilty (X_2). If guilty pleas still go down, the number of cases going to trial does not have to go up (X_3) if the prosecutors are willing to dismiss more cases that are of low importance in the criminal justice system. If the number of trials still goes up, delay (X_4) can be held constant by hiring more judges. If delay still goes up, the jail population can still be reduced (-Y) by giving priority trials to those who are in jail well ahead of those who are out of jail. Thus, good models like those which relate to bilateral bargaining, queueing backlogs, queueing delay, and population growth can be used both for predicting consequences and also for prescribing counter-policies.[4]

D. THE EQUITY DILEMMA

The equity dilemma refers to the frequent conflict in policy evaluation between policy goals of efficiency and equity. Efficiency refers to choosing policies that will maximize benefits minus costs. Sometimes the term is used to refer to choosing policies that will provide the highest benefit/cost ratio. The two definitions produce identical results when costs are held constant. When they are not held constant, the first definition will result in policies that will produce more net worth or net satisfaction provided that a minimum benefit level is met and a maximum cost level is not exceeded. Equity is considered to be present when there is equality per capita across all groups or places with regard to either receiving benefits or paying costs. Instead of equality per capita, equity may refer to equality of benefits per needs of the recipients when there are substantially different needs, and equity may refer to equality of costs per ability to pay when there is substantially

different ability to pay on the part of those paying the costs. There are many ways of measuring equity or equality. Perhaps the simplest is to determine the average deviation from the average benefit score or the average cost score. The closer the average deviation is to zero, the more equity is present, assuming equal needs and/or equal ability to pay.

A good illustration of the efficiency versus equity dilemma is in the area of allocating federal anti-crime dollars across cities or states. If the money were to be allocated just on efficiency grounds, we might determine for each city the slope or marginal rate of return between crime occurrence and anti-crime dollars. We could do that by obtaining data showing for many previous points in time how much crime occurred (Y) and how many anti-crime dollars were spend (X) in order to obtain a nonlinear regression equation for each city of the form $Y_t = a + b_1(\log X_{t-1}) + b_2(Y_{t-1})$. The X or dollars variable is lagged one time-period because it takes awhile for anti-crime dollars to have an effect on crime, and because we want to avoid the reciprocal effect of crime on increasing anti-crime dollars. The X or dollars variable is also logged to reflect the fact that although anti-crime dollars may decrease crime, they do so at a diminishing or logarithmic rate. Crime at a prior point in time is included as a control variable in order to hold constant indirectly all the variables that cause crime other than anti-crime dollars. With that kind of equation for each city, we could then allocate a federal budget to the cities in accordance with their negative b_1 regression coefficients. One can algebraically prove that doing so will result in minimizing predicted crime occurrence across the cities, assuming that the regression coefficients are accurate. In other words, that type of allocation will provide maximum national benefits at a given budget and thus be maximally efficient.

Such an efficient allocation, however, may be quite inequitable. Perfect equity or equality would not, though, involve allocating the budget across the cities on a purely per capita basis. That would only equalize them on the anti-crime dollar-benefits received. Some cities would be bearing higher costs than others, not just in the sense of paying more taxes per capita, but in the more relevant sense in this context of suffering more crime per capita. An alternative equity allocation would involve trying to give to each city an amount such that every city would have the same quantity of crime. Even if this were politically, economically, and algebraically possible to do, we would still have a trilemma

among allocating on the basis of (1) efficiency, (2) equality of dollars, and (3) equality of crime.

The solution to this problem might be to allocate in three steps as follows:

1. Give to each city a minimum allocation on the basis of population in order to satisfy dollar equity.
2. Give to each city whose per capita crime-level exceeds a maximum threshold an amount needed to bring them down to that maximum threshold.
3. Then, with what is left over after steps 1 and 2, allocate the remainder in accordance with the regression or efficiency coefficients.

This approach may have to be considerably modified due to the inability to get meaningful negative regression coefficients. Instead, one may have to do step 2 in light of a somewhat arbitrary formula that gives extra weight to cities with a bad crime problem. Likewise, step 3 may also require a formula that allocates to the cities in accordance with subjective and objective judgments concerning the efficiency of the personnel, fiscal, organizational, and other management aspects of their police, courts, and corrections systems.

The important point is not the technical aspects of how one considers equity of benefits, equity of costs, and efficiency, but rather that those criteria are consciously considered. Policy evaluators may not have an obligation to arrive at a solution that is both efficient and equitable, which may be impossible. They should, however, have an obligation to explicity discuss these issues in their policy evaluations. Efficiency is clearly important by definition if policy evaluators are trying to maximize benefits minus costs. Equity is important partly for political reasons in the sense that if any group or place feels it is receiving benefits below a minimum level or suffering costs above a maximum level, then that group or place may rebel against the government at the ballot box or possibly even at the barricades. Equity is also important for public image reasons in the sense that governments are embarrassed internationally and domestically by certain groups or places suffering to extremes. Equity can also be justified on more philosophical grounds that draw upon notions of inherent right found in the writings of Aristotle, Aquinas, Locke, Rousseau, and Marx, as well as the basic books of all major religions.[5]

II. ETHICAL DILEMMAS THAT RELATE TO APPROPRIATE METHODS

A. THE SENSITIVITY ANALYSIS DILEMMA

In doing policy evaluation, to what extent do policy evaluators have an obligation to show how their predictive or prescriptive conclusions would be changed as a result of changes in the data, measurement, sampling, assumptions, values, or other inputs to which the conclusions might be sensitive? This is a dilemma in the sense that, on the one hand, one has limited resources for conducting elaborate research, and we would not want to deter people from doing non-elaborate research that might be useful. On the other hand, we do not want policy makers, the public, and others to conclude falsely that policy findings are more widely applicable than they might be. This dilemma, like the optimizing one, can often be resolved by researchers at least raising questions about the sensitivity of the findings to changes in the inputs, even if they do not test those sensitivity hypotheses.

Doing an elaborate sensitivity analysis is no guarantee that it will be referred to by policy makers, but researchers will at least have performed their roles properly. For example, an elaborate analysis was done on the question of an optimum jury size that the U. S. Supreme Court referred to in a key case dealing with the constitutionality of alternative jury sizes. The research showed that with a number of assumptions, the optimum jury size would be approximately 7. The Supreme Court by way of Justice Blackmun's opinion found that juries smaller than 6 were unconstitutional, but they could be smaller than 12, partly citing research on optimum jury size. That research, however, showed that an optimum jury size is dependent on (1) the relative tradeoff value (V) of convicting an innocent defendant versus not convicting a guilty defendant, (2) the percentage (P) of innocent defendants and guilty defendants who might be convicted at various jury sizes, and (3) the number of defendants (N) in jury trials who are truly innocent and the complementary percentage of who are truly guilty per 100 defendants.

The tradeoff value (between the two types of errors) that Justice Blackmun accepted was 10 to 1 in view of its advocacy by William Blackstone who was considered to have been influential on the framers of the Constitution. That normative input was thus not so subject to change. On the empirical input of the conviction percentages, one can show the

optimum jury size is not substantially influenced by reasonable values assigned to that input. This is true partly because the average defendant is known to be convicted at a 64% rate, and therefore innocent defendants are probably convicted at less than a 64% rate, and guilty defendants at more than a 64% rate. Justice Blackmun, however, did choose to ignore the conclusion's sensitivity to the third input, which is a mixture of empirical and normative elements. For the sake of discussion, it was assumed that prosecutors operate at a 95% confidence level in bringing defendants before jury trials such that they feel 95% confident the defendants are truly guilty. If, however, one thinks prosecutors are not so careful and are operating at only the 90% confidence level, then much bigger juries deciding by a unanimous vote are needed to protect those additional innocent defendants from being convicted. On the other hand, if one thinks prosecutors are especially careful and are operating at a 99% level, then we could have much smaller juries since there would be fewer innocent defendants to protect from being convicted. The correct confidence level and (more important) the correct number of truly innocent defendants who are tried per 100 defendants is impossible to determine. Thus, the bottom-line prescription is not only dependent on a highly sensitive input, but one that is also impossible to measure.

The optimum jury size between 6 and 12 is thus ultimately dependent on whether one has a liberal pro-defense perception of the criminal justice system (for which 12 would be the optimum), or a conservative pro-prosecution perception of the criminal justice system (for which 6 would be optimum). Researchers in this context have done all they can in pointing out those sensitivity analysis matters, but they have not performed so properly if those matters are ignored or neglected.[6]

B. THE EFFICIENT RESEARCH DILEMMA

The efficiency-equity dilemma relates to societal efficiency and equity. The efficient-research dilemma relates to the matter of doing evaluation research that does not waste money on an evaluation that could be done about as well for substantially less. In other words, policy evaluators have an ethical obligation to try to provide maximum benefits minus costs to society or their clients with regard to both their research projects and their policy recommendations. The dilemma occurs by virtue of the fact that truly efficient research may result in depriving the researcher of additional salary, released academic time,

interesting travel, pretty or handsome secretaries, and data that might have future personal value. It also occurs since we do not want to encourage evaluators to be unduly research-cost conscious and thereby fail to make expenditures that would provide worthwhile research benefits. If increasing the research budget will truly benefit both the interests being served and the researcher, then there is no dilemma.

Many examples can be given of evaluation research projects that may have been overly expensive relative to the insights obtained, partly because the budget was inflated in covering salaries, equipment, travel, and other items. The personnel from virtually all funding units and research units can tell about many such instances. What may be needed is more simplicity in policy evaluation, not only to save money, but also to facilitate more policy evaluation being done and being more meaningfully communicated to policy makers. Perhaps a good example of simplicity in policy evaluation might be the analysis of alternative ways of providing attorneys to the poor. That problem basically involves three policy choices. The non-criminal legal problems of the poor can be handled mainly by either (1) volunteer attorneys, (2) salaried government attorneys, or (3) private attorneys who are reimbursed by the government for representing the poor. The problem basically involves four policy goals, namely, having a program that is (a) inexpensive, (b) visible and accessible, (c) politically feasible, and (d) likely to be staffed by lawyers who are technically competent concerning poverty law and reasonably aggressive. On the goal of inexpensiveness, volunteer attorneys clearly score the best. Private attorneys who are reimbursed by the government are so expensive as to make that approach economically unfeasible, and thus not worth discussing further. It was endorsed by President Reagan before he was elected as being a private marketplace solution, but rejected afterwards in favor of volunteer attorneys because of the expense. Salaried government attorneys are clearly more visible and accessible than volunteer attorneys. The volunteer system, however, is much more politically feasible since it is not so likely to generate a cadre of lawyers who are disturbing to landlords, merchants, employers, welfare bureaucrats, and local politicians as the salaried legal services program is sometimes viewed. On the other hand, salaried government attorneys are more likely to develop specialized competence and be more willing to go to court than volunteer

attorneys, who tend to be inexperienced new attorneys looking for experience.

Thus, which is the better program between volunteer attorneys and salaried government attorneys depends on the relative weight one gives to the four goals. Conservative values are likely to weight inexpensiveness and political feasibility high. Liberal values are likely to weight visibility/accessibility and competence/aggressiveness high. Thus, with conservative value-weights, the volunteer system is best. With liberal value-weights, the salaried attorney system is best. That type of analysis may lead one to recommend a compromise between the two systems in order to obtain support from both conservatives and liberals. A compromise might involve salaried government lawyers who have as one of their main functions to coordinate the activities of volunteer lawyers, especially with a system that requires lawyers to do some volunteering as part of their license to practice law.

That type of analysis can be done on the back of an envelope at zero cost assuming the envelope were found in a wastebasket and the evaluator were commuting to work and thus not on governmental time. The cost could be increased by doing a literature search to build on what is already known, and by interviewing people knowledgeable about alternative ways of providing legal services to the poor. The Legal Services Corporation spent over $10,000,000 to analyze the effects of alternative delivery systems. Much of that money went to pay for setting up programs that could be observed, rather than observing already existing programs. Much of the money also went to expensive data-gathering activities which have yet to be made accessible to other researchers. It is questionable whether the highly incremental costs over zero costs were justified by the not-so-highly incremental benefits over what is common sense to knowledgeable insiders. There is a need in policy evaluation for making more use of knowledgeable insiders and especially for emphasizing more thought and creativity, rather than expensive data gathering, if we are going to have more efficient and useful policy research.[7]

C. THE RESEARCH SHARING DILEMMA

One of the most frustrating aspects of policy evaluation is not the difficulty of gathering data that is needed to analyze alternative public policies, but rather the difficulty one sometimes encounters in getting access to relevant data that has already been

gathered. The research sharing dilemma involves an evaluation researcher being faced with the choice between withholding or sharing expensively gathered data. On the one hand, the researcher might like to keep the data to avoid embarrassing himself or others, and to be able to use it for future research without others doing so. On the other hand, if we are talking about societal desirability as a dilemma-resolving criterion, then in almost all these instances, the data should be shared so as to facilitate the development of new knowleege on which scientific research thrives, including evaluation research.

Many examples could be given of data that was withheld which subsequently proved useful for secondary analysis of important policy problems. One example is the American Bar Foundation data set covering 12,000 criminal cases randomly sampled from all 50 states, It was gathered in 1963 at great expense by thousands of lawyers from their county courthouses in accordance with detailed instructions. Lee Silverstein of the American Bar Foundation was responsible for coordinating the data gathering, mainly for the purpose of determining the extent to which attorneys were representing criminal defendants and the effects of having an attorney. After authoring a landmark book on Defense of the Poor in Criminal Cases in American State Courts (Little, Brown, 1965), Lee Silverstein died. The original data and the codebooks were in the process of being thrown away when they were rescued for inclusion in the Interuniversity Consortium for Political Research at Ann Arbor, Michigan. Since then, that data set has been reused many times to test hypotheses relating to such matters as (1) the extent to which there are disparities in the criminal justice system for various stages, crimes, and types of defendants, (2) the occurrence of delay at various stages and how delay relates to a variety of variables, and (3) various aspects of plea bargaining, pretrial release, sentencing, and jury decision-making, as well as aspects of right to counsel that Lee Silverstein did not cover. This is an example of data that was almost withheld or destroyed due to an innocent or negligent lack of understanding as to how useful such data can be for other research.

A frequent motivation for withholding data is that someone in the sample might be embarrassed by its release. That is normally not a good justification because identifiers can usually be removed so that there is no way of tracing the machine-readable data to any specific respondents or persons. One example is the OEO Legal Services Program's data-set

covering about 200 variables for each of about 250 individual legal services agencies across the country. The data was originally gathered in 1970 to determine which agencies should get increased, decreased, or constant funding, and would have possibly been destroyed after serving that purpose. The data was, however, pried loose after considerable negotiation. Since then, the data has been used to test hypotheses as to what variables correlate with having high scores on a variety of ratings among legal services agencies. The data has also been used to illustrate alternative ways of allocating scarce resources across activities and places.

The most frustrating motivation for withholding data is a profit-seeking desire to cultivate the data in future contracts with government funding agencies. That has been known to occur with research institutes that have received large amounts of money from the Law Enforcement Assistance Administration to gather and analyze data. Their excuses tend to emphasize that the data is not ready for distribution, no matter how much they have already used the data. To the credit of LEAA, much of that data has now been archived at the Interuniversity Consortium. LEAA has succeeded in prying it loose from the research institutes that were paid with tax money to gather the data, but those institutes tend to be rather uncooperative in explaining aspects of their codes to secondary researchers. This illustrates that perhaps the best way to get ethical behavior out of evaluation researchers is to make the costs of unethical behavior substantially greater than the benefits, although that in a general way may be the best way to encourage any kind of socially desired behavior.[8]

D. THE RESEARCH VALIDITY DILEMMA

Research validity refers to having conclusions that (1) conform to empirical reality and that (2) logically follow from meaningful premises. Policy researchers may have a greater obligation to be concerned with research validity because their conclusions are more likely to have societal consequences than research that is more abstract and less policy-relevant. On the other hand, policy researchers often need not be as concerned with precise measurement because policy problems often merely involve determining which alternative policy is best in light of given goals, not the exact position of all the alternative policies. Guaranteeing research validity may also sometimes require extra effort and cost, which runs contrary to the cost-saving obligation of policy research. If, however, there is a conflict

between research validity and cost saving, validity would probably deserve a higher priority, although that depends on how much extra validity one would obtain for the extra cost.

Empirical validity tends to depend on (1) the representativeness of the sample of persons, places, or things on which the hypotheses are tested, (2) the accuracy of the measurement of the characteristics or variables of those entities, and (3) the way in which that data is analyzed, particularly with regard to considering other variables that might explain the relations between the main variables. A good example to illustrate invalidity in policy-relevant research along these lines might be the subject of whether black or poor defendants are treated worse than white or non-poor defendants in the criminal justice system. On the matter of representativeness of the entities sampled, most of the literature on the subject tends to involve a single city at a single point in time. That is fine for building knowledge by way of accumulating such case studies. Individual researchers, however, sometimes have a tendency to treat their case studies as indicative of more general relations even though a northern city may be very different from a southern city, and a rural area very different from an urban area. More important, however, may be the fact that most of the studies concentrate only on the sentencing stage in the criminal justice process, but yet tend to talk about racial disparities on a more general level. In reality, the sentencing stage may be the least discriminatory because by then, the defendants have been filtered down to being more homogeneous. This is in contrast to the earlier stages of who gets arrested for similar crimes, who gets released on bail, who has a hired lawyer, and who has a jury trial.

On the matter of measurement, it is common to score probation as a zero sentence on the sentencing variable, rather than have two sentencing variables, one on probation versus jail and the other on length of jail sentence. By lumping both decisions together, offsetting discriminatory patterns can disappear. That is so if whites do better on probation versus jail possibly due to racial considerations, but blacks do better on getting shorter jail sentences possibly by virtue of having committed less serious crimes, especially less serious larceny. On the matter of analysis and controlling for confounding variables, many studies attempt to control for the crime committed by looking to its average or maximum sentence. It might be more meaningful to look to whether the crime is one against persons or property. By lumping more severe crimes together, regardless whether they involve assault or larceny,

offsetting discriminatory patterns can also disappear. That is so if whites do better than blacks in larceny cases, which are often black-on-white; but blacks do better in assault cases, which are relatively more likely to be black-on-black.

On the matter of drawing conclusions that logically follow from meaningful premises, the most illogical reasoning may involve the traditional decision rule that if the differences found can be attributed to chance at more than a .05 probability level, then we will act as if the differences are chance differences rather than real differences. That .05 premise implies that it is 19 times as bad to falsely accuse the system of being discriminatory than it is to falsely exonerate the system as being nondiscriminatory. It would be hard to justify that 19 to 1 trade-off value in terms of the relative benefits and costs of these Type 1 and Type 2 errors and accuracies. Likewise it may be quite illogical to say that a relation of .10 between a dependent variable and an independent variable will not be considered substantial, since squaring that relation indicates it only explains 1% of the variance on the dependent variable. Such a relation may be very substantial when the .10 is the percentage-points difference between convicted blacks receiving jail sentences (rather than probation) versus convicted whites. This is especially so if 0% of the whites go to jail and 10% of the blacks do so, since that means being black is a necessary condition for going to jail. The .10 difference is also especially important if 90% of the whites go to jail and 100% of the blacks do so, since that means being black by itself is a sufficient condition for going to jail. A .10 difference where jail terms are at stake can create crime-causing anti-social attitudes, and it is thus something to be concerned about and not passed off because a non-meaningful premise says squared regression coefficients have to be greater than 1%.

Policy evaluation is basically deductive analysis from inductively arrived at premises, in the sense that policy analysis can be simplified to a syllogism of the form (1) Y is good, (2) X causes Y, and (3) therefore adopt X. Determining whether Y is good, or whether X causes Y may involve (1) surveying the values or perceptions of relevant people, (2) statistically relating Y to a higher goal, or X to Y, (3) deducing the premise

from some other premises, or (4) assuming the premise is true subject to a sensitivity analysis to show how the conclusions would change for alternative premises. Those methods potentially involve empirical and logical validity problems to which policy researchers need to be sensitive.[9]

E. THE DILEMMA OF PUTTING PEOPLE AT RISK

One of the most controversial ethical issues in policy evaluation is the extent to which experiments should be performed to test the effects of alternative public policies where the experiments involve putting people into situations where they risk being physically, mentally, or otherwise harmed. The dilemma is one of wanting to validly know the effects of a proposed policy versus not wanting to harm people who otherwise would not be harmed if it were not for the experiment. In recent years, national policy experiments have been performed where people were randomly assigned to experimental groups and control groups. Those experiments have included giving welfare recipients different amounts of aid to see how the differences would affect their ambition to get jobs,or giving some ex-convicts unemployment compensation and denying it to others to see how receiving it would affect their crime committing. The welfare experiment did not involve putting people at serious risk since no one was included in a category that would jeopardize one's health, although the experiment may not have been so successful in demonstrating the effects of payment differences given the small range of differences. Denying unemployment compensation to some ex-convicts was not a significant ethical problem because recently released convicts normally do not get unemployment compensation.

An example of an important criminal justice experiment that does raise serious ethical problems is the experiment which randomly assigned some juvenile delinquents to jail and some to probation. In 1966, convicted juvenile delinquents in Sacramento and Stockton were divided into three groups as follows:

1. Those who committed minor crimes and who had minor prior records. They received probation.
2. Those who committed major crimes and/or who had major prior records. They

received jail sentences.

3. Those who fit into a middle third of relatively borderline cases in terms of their offenses and their prior records. Half of them were randomly assigned to one-year probation terms and half to one-year jail terms.

Then a year after the completion of the probation and jail terms, the crime-committing records of both groups were compared. The probation group was found to have a better post-probation record than the post-jail record of the jail group. That could not be attributed to any differences in their pre-conviction records since both groups had similar pre-conviction records given the random assignment to being in the probation or jail groups. One can argue that the experiment was not unethical because those who were included could have justifiably received either probation or one-year sentences in view of their borderline characteristics. On the other hand, minimum due process in sentencing would require something more than a flip of a coin to determine who goes to jail and who stays out, even if useful scientific knowledge were at stake. Subsequent to that experiment, the Supreme Court said juvenile delinquents were entitled to basic due process rights (in the 1967 case of In re Gault), which probably would have made this experiment unconstitutional.

Although it is important to be sensitive to not putting ordinary individuals at risk of harm without considerable offsetting justification, almost the opposite presumption might be present when the individuals are public officials. In that context, they should be protected from being physically harmed or having irrelevant aspects of their personal lives exposed. When, however, it comes to subjecting them to embarrassment for making extreme, illegal, or otherwise objectively wrong decisions, then policy researchers have an obligation to expose that kind of wrongdoing, rather than shelter those subjects from risk. A concrete relevant example involved a proposed research project in which data from approximately ten judicial districts was to be gathered, showing for each judge the percentage of defendants released and the percentage of released defendants who appeared for trial. It was thought that judges would differ greatly on their release rates, but very little on their appearance rates. The experiment involved randomly assigning some districts

to an experimental group in which those rates were to be publicized among the judges, the bar, and possibly the press, and assigning some districts to a control group in which the rates would not be publicized. In the after-period, data was to be collected to see how the publicity affected changes in the release rates. It was anticipated that where the publicity was the strongest, the high-holding judges would be embarrassed into bringing down their holding rates in view of the extra costs they were incurring without offsetting benefits. The research was subjected to considerable bureaucratic red tape because it used human subjects who were put at risk, and the research was never funded partly because it was considered too controversial. Thus, the dilemma of putting people at risk involves a desire to avoid harming people on the one hand, and a desire for validity on the other hand, and sometimes also a desire to test hypotheses relevant to determining and decreasing socially undesirable behavior by public officials.[10]

III. SOME CONCLUSIONS

Looking over the discussion of these nine ethical dilemmas, one is possibly struck by two underlying factors. One factor relates to how these dilemmas differ from the ethical dilemmas in traditional social science research, and the other relates to a key factor that seems relevant to helping resolve the dilemmas.

A. POLICY EVALUATION VERSUS TRADITIONAL SOCIAL SCIENCE

In policy evaluation, there is a concern for arriving at policies that will achieve or maximize societal or group goals. In traditional social science research, there is a concern for establishing reality particularly about predictive and causal relations, regardless whether the findings have any policy or practical usefulness. Thus, there is generally more at stake in each of these dilemmas than there are in the counterpart situations in traditional social science research. In the average social science study, one normally does not care relatively so much about the nine counterpart situations as to:

(Counterpart situations that relate to purposes, goals, and effects)

1. Whether an optimum relational model has been developed.

2. Whether a political party or interest group uses or misuses the findings.
3. Whether unanticipated relations are found when analyzing data.
4. Whether one studies income equality rather than worker productivity.

(Counterpart situations that relate to methods)

5. Whether the conclusions are sensitive to changes in the sample, measurement, or analysis.
6. Whether the research costs are doubled if they are relatively small to begin with.
7. Whether data is shared if it has relatively little practical value anyhow.
8. Whether the research is valid if the consequences of invalidity are relatively slight.

On the ninth dilemma of putting people at risk of harm, traditional social scientists are more concerned than policy evaluators, partly because social scientists can more often manipulate people experimentally, and because they may find it more difficult to offer a socially valuable justification for doing so. The above points are obviously rough generalizations since there are plenty of important social science studies in which more may be at stake than plenty of unimportant policy evaluation studies. In general, though, policy evaluators do seem to need a higher sense of responsibility in performing their professional functions than ordinary social scientists do, given the greater potential societal benefits and costs that tend to be involved.[11]

B. OPENNESS IN RESOLVING ETHICAL DILEMMAS

Openness seems to be the key factor underlying the resolving of the ethical dilemmas in policy evaluation. The situation is like blue sky laws in regulating the securities markets. A stockbroker can legitimately sell the blue sky so long as he informs potential customers and others that is what they are getting. Likewise policy evaluators are complying with basic ethical obligations when they make clear that they are:

(Openness in relation to purposes, goals, and effects)

1. Merely determining the relation between a given policy and a given goal, rather than seeking to find a policy or combination that will maximize a given goal or alternative goals.

2. Working for a partisan cause or interest group, rather than giving the illusion that one is performing in a non-partisan scholarly role.
3. Checking for only certain consequences so the evaluation consumer can at least see what consequences have not been checked for.
4. Considering equity and efficiency in certain ways among various available alternatives, so evaluation consumers can decide whether or not they agree.

(Openness in relation to methods)

5. Merely drawing a conclusion on the basis of a single set of measurements, entities, assumptions, or other inputs, rather than seeking to vary these inputs to determine how such variation would affect the conclusions.
6. Incurring certain itemized costs as part of the research project, and offering as much justification as seems reasonable.
7. Making available their data for secondary analysis, but without necessarily doing anything expensive in cleaning or documenting the data so long as potential users are informed of possible defects in the data or the documentation.
8. Working with certain measurements, samples, assumptions, and other evaluation inputs which allow for certain conclusions and no others.
9. Openly informing people who might be part of an experiment as to what risks they are being subjected to, so they can opt out of the experiment if they want to do so, unless there are important offsetting considerations such as experiments to determine or decrease wrongdoing on the part of public officials.

C. AFFIRMATIVE ACTION IN RESOLVING ETHICAL DILEMMAS

Perhaps more affirmative action is needed for resolving these ethical dilemmas than just openness in reporting one's research. In the context of the blue sky laws, one might argue that a stockbroker should have an ethical obligation not to sell the blue sky, rather than merely an ethical obligation to report that he or she is selling the blue sky. On the above nine dilemmas in policy analysis research, the more affirmative position might involve the following:

(Affirmative action in relation to purposes, goals, and effects)

1. Policy Analysts should seek to find the policy or combinations of policies that will maximize a given goal or goals.
2. They should not aid certain groups if the goals of the groups are illegal, such as racist redistricting designed to decrease the influence of minority groups below a reasonably proportionate level.
3. They should check for side effects and unforeseen consequences in doing policy analysis.
4. They should be concerned with equity as a minimum constraint, and also as a constraint that is subject to a sensitivity analysis for determining how the prescriptions would differ if the equity constraints were changed.

(Affirmative action in relation to methods)

5. They should always do a sensitivity analysis to determine how their prescriptions might differ as a result of changing the measurement, entities, assumptions, or other inputs of their basic analysis.
6. They should try to keep research costs down.
7. They should seek to facilitate the secondary re-use of data which they have gathered.
8. They should be careful to avoid findings of questionable validity, including failure to control for alternative explanations or failure to use proper measurements, samples, and assumptions.
9. Policy analysts should be careful not to put people at risk, or else be sure informed consent is involved if people are put at risk, although policymakers can and should be subject to being exposed as wrongdoers or as having undesirable biases in their decisionmaking.

On the basis of this analysis one can possibly arrive at an overall conclusion that policy evaluators operating on a high level of professional responsibility should seek to develop optimum policies and to do optimum research. Optimum policies refer to policies that maximize societal benefits minus societal costs. Optimum research refers

to research that maximizes beneficial new knowledge minus the costs of obtaining it. Those are high and often impossible goals to achieve. By seeking them, however, policy evaluation is likely to achieve more than it would with any lesser set of goals.

FOOTNOTES

1. On the ethics of policy evaluation in general, see Gordon Bermant, Herbert Kelman, and Donald Warwick (eds.), The Ethics of Social Intervention (Wiley, 1978); Daniel Callahan, The Teaching of Ethics in Higher Education (Hastings Center, 1982); William Dunn (ed.), Ethics, Values, and the Practice of Policy Analysis (Lexington-Heath, 1982); Joel Fleishman and Bruce Payne, Ethical Dilemmas and the Education of Policymakers (Hastings Center, 1980); Keith Marvin, Standards for Program Evaluation (Evaluation Research Society, 1980); and John Rohr, Ethics for Bureaucrats: An Essay on Law and Values (Dekker, 1978).

2. On an optimizing perspective to policy evaluation, see Duncan MacRae and James Wilde, Policy Analysis for Public Decisions (Duxbury, 1979); Edith Stokey and Richard Zeckhauser, A Primer for Policy Analysis (Norton, 1978); and S. Nagel, Policy Evaluation: Making Optimum Decisions (Praeger, 1982). On a relational perspective to policy evaluation, see Thomas Cook and Donald Campbell, Quasi-Experimentation: Design and Analysis Issues for Field Settings (Rand McNally, 1979); Laura Langbein, Discovering Whether Programs Work: A Guide to Statistical Methods for Program Evaluation (Goodyear, 1980); and David Nachmias, Public Policy Evaluation: Approaches and Methods (St. Martin's, 1979). On the example of relating causal factors to crime versus making resource-allocation recommendations, see "Tradeoffs in Crime Reduction Among Certainty, Severity, and Crime Benefits" in S. Nagel, The Legal Process and Policy Evaluation (Unpublished chapter available on request, 1982).

3. On the ethics and legality of working with political parties and interest groups, see Daniel Gaby and Merle Treusch, Election Campaign Handbook (Prentice-Hall, 1976), 289-296; and Harmon Zeigler and Michael Baer, Lobbying: Interaction and Influence in American State Legislatures (Wadsworth, 1969), 60-80. On the example of working with political parties and interest groups in the field of legislative redistricting, see S. Nagel, "Computers and the Law and Politics of Redistricting," 5 Polity 77-93 (1971).

4. On unforeseen consequences in policy evaluation, see Sam Sieber, Fatal Remedies: The Ironies of Social Intervention (Plenum, 1981); Harry Jones, The Efficacy of Law (Northwestern, 1969); and Helen Ingram and Dean Mann, Why Policies Succeed or Fail (Sage, 1980). On the example of the relation between increasing pretrial release and thereby increasing the jail population, see S. Nagel and M. Neef, "Department of Unintended Consequences: Two Examples from the Legal Process," 2 Policy Analysis 356-360 (1976).

5. On equity as a goal in policy evaluation, see Edmund Phelps (ed.), Economic Justice: Selected Readings (Penguin, 1973); and John Rawls, A Theory of Justice (Harvard, 1971). On efficiency as a goal in policy evaluation, see Roland McKean, Efficiency in Government Through Systems Analysis (Wiley, 1966); and Werner Hirsch, The Economics of State and Local Government (McGraw-Hill, 1970). On the conflict between the two, see Arthur Okun, Equality and Efficiency: The Big Tradeoff (Brookings, 1975). On the example of allocating anti-crime dollars across places, see "Allocating Anticrime Dollars across Places and Activities" in S. Nagel, Policy Evaluation: Making Optimum Decisions (Praeger, 1982).

6. On sensitivity and post-optimizing analysis, see David Anderson, Dennis Sweeney, and Thomas Williams, An Introduction to Management Science: Quantitative Approaches to Decision-Making (West, 1976), pp. 143-170; Hamdy Taha, Operations Research: An Introduction (Macmillan, 1971), pp. 94-109; and Harvey Wagner, Principles of Operations Research: With Applications to Managerial Decisions (Prentice-Hall, 1969), pp. 129-164. On the example of the sensitivity of an optimum jury size to various inputs, see S. Nagel, "Management Science and Jury Size," 11 Interfaces: A TIMS-ORSA Journal 34-39 (1981).

7. On cost saving in evaluation research, see "No Money, No Time, No Technique," in James Glass and Charldean Newell, Administrative Research Methods (Unpublished manuscript, 1982); and Rick Hesse and Gene Woolsey, Applied Management Science: A Quick and Dirty Approach (Science Research Associates, 1980). On the example of evaluating alternative ways of providing legal services to the poor, see S. Nagel, "How to Provide Legal Counsel for the Poor: Decision Theory," in Dorothy James (ed.), Analyzing Poverty Policy

(Lexington-Heath, 1975); and Legal Services Corporation, The Delivery Systems Study (LSC, 1980).

8. On research-data sharing, see Herbert Hyman, Secondary Analysis of Sample Surveys (Wiley, 1972); Ralph Bisco, "Social Science Data Archives: A Review of Development," 60 American Political Science Review 93-109 (1966); and Inter-University Consortium for Political and Social Research, Guide to Resources and Services (ICPSR, 1978).

9. On research validity, see Abraham Kaplan, The Conduct of Inquiry: Methodology for Behavioral Science (Chandler, 1963); and Eugene Meehan, The Theory and Method of Political Analysis (Dorsey, 1964). On the example of research validity in the context of black-white differences in sentencing, see "The Racial Disparity That Supposedly Was Not There: Some Pitfalls to Watch for in Analyzing Court Statistics on Sentencing and Other Matters" in S. Nagel, The Legal Process: Modeling the System (Sage, 1977).

10. On putting people at risk in social science experiments, see Federal Judicial Center, Experimentation in the Law (FJC, 1981); Jay Katz, Experimentation with Human Beings (Russell Sage, 1972); and Paul Freund (ed.), Experimentation with Human Subjects (Braziller, 1970). On the example of randomly assigning convicted juvenile delinquents to jail or probation, see Martin Levin, "Policy Evaluation and Recidivism," 6 Law and Society Review 17-46 (1971). On the example of randomly assigning judicial districts to being subjected to different kinds of publicity concerning the holding and appearance rates of individual judges, see S. Nagel, "Decision Theory and Pretrial Release Decisions" (Unpublished research proposal available from author on request, 1977).

11. On ethics in general political and social science research, see "Ethical Concerns in Social Science Research" in David and Chava Nachmias, Research Methods in the Social Sciences (St. Martin's, 1981); "The Ethics of Political and Social Inquiry," in Dickinson McGaw and George Watson, Social and Political Inquiry (Wiley, 1976); and Nicholas Hobbs, "Ethical Issues in the Social Sciences," 5 International Encyclopedia of the Social Sciences 160-166 (1968).

12. On openness as a way of controlling wrongdoing in securities regulation and

governmental activities, see F. B. Ashby, Economic Effects of Blue Sky Laws (U. of Pennsylvania Press, 1926); Louis Loss, Securities Regulation (Little, Brown, 1951); Harold Cross, The People's Right to Know (Columbia, 1953); and Charles Whalen, Your Right to Know (Random House, 1973).